The inefficacy of superficial similarities for improving instructor-student relationships*

Replicating 'Taking It to the Next Level'

Amy Farrow

April 26th, 2021

Abstract

This paper replicates the 2019 article 'Taking It to the Next Level' (Robinson, Scott, and Gottfried 2019), which evaluates an intervention to improve college instructor-student relationships. The modeling results indicate that the intervention, consisting of informing instructors and students about commonalities, has a weak positive effect on student perceptions of instructor-student similarity, but no effect on student perceptions of instructor-student relationship, instructor perception of similarity, or instructor-student relationship. The goal of this intervention was to improve student performance and persistence, and while the scalability and affordability are desirable, there are no measurable effects on any of the student outcome measures, including grades and reenrollment. These results are consistent with the original paper.

Keywords: replication study, instructor-student relationships, college, similarity, college persistence

Contents

*Code and data are available at: github.com/amycfarrow/takingittothenextlevelrepro.

1 Introduction

Despite the growing importance of college degrees for career success, college completion rates are a continuing problem. The causes are complex: factors that affect student persistence include student characteristics, organizational factors, peer environment, and student experiences. Interventions to improve student persistence can target any of these aspects.

Robinson, Scott, and Gottfried (2019)'s paper, "Taking It to the Next Level: A Field Experiment to Improve Instructor-Student Relationships in College", tests an intervention to improve college student persistence and performance. In this field experiment, they evaluated the links between perceived similarity, instructor-student relationships (ISRs), and student success. Based on extensive K-12 research about the importance of instructor-student relationship for student success, Robinson, Scott, and Gottfried (2019) aimed to establish how instructor-student relationships could be improved at the college level, and to test if this improvement had a positive impact on measurable student outcomes. The experiment consisted of a randomized controlled trial where some undergraduate students were informed of similarities they shared with their instructor, while others were not. Student and instructor perceptions of similarity and the instructor-student relationship were measured through surveys, and student performance measures were collected from school records.

This paper replicates Robinson, Scott, and Gottfried (2019)'s original analysis, using anonymized data provided by the authors. First, Robinson, Scott, and Gottfried (2019)'s methods are discussed, and the measures and scales are evaluated. Second, demographics of instructor and student populations are explored, similarity of treatment and control groups is established, outcomes of interest are compared for treatment and control groups, and missing data is evaluated. Third, Robinson, Scott, and Gottfried (2019)'s models are replicated, and additional models are explored. Linear, logistic, and ordinal logistic models are replicated to predict outcomes including ISR perception, grade, and persistence. Further replicated exploratory models consider the relationship between similarity and instructor-student relationship, as well as between instructor-student relationship and student outcomes. Additional models use different controls–specifically, variables indicating shared racial and gender self-identification.

These models show that treatment slightly improves student perception of instructor-student similarity, but does not significantly affect student or instructor ISR perception, grades, or persistence. These conclusions stand when matching instructor-student traits are considered and when vulnerable student groups are evaluated separately from the wider student population. The lack of significant effects reflects the complicated multi-stage task of increasing perceived similarity, to improve instructor relationships, and to then improve student outcomes, especially in a college context that may make instructor-student similarity and connection challenging.

1.1 Literature review

College is perceived to be a meritocratic tool for social mobility and career success (Bowen, Chingos, and McPherson 2009). Unfortunately, retention is a large problem in contemporary American colleges, and many students begin degrees without completing them; six-year completion rates for full-time first-time students range from 51% to 86%, depending on school (Delbanco 2015). Even when controlling for pre-college test scores and initial enrollments, completion rate disparities exist based on parental education, socio-economic status, and race/ethnicity (Delbanco 2015). Disparities in college completion lead to further entrenchment of long-standing inequalities (Delbanco 2015). Thus, measures to help students persist in college completion are desirable to reduce wasted resources and to increase societal equity.

Many causes, and corresponding solutions, have been theorized for the ‘college completion crisis’. In his examination of persistence research, Reason (2009) considers student precollege characteristics (sociodemographic traits, academic preparation and performance, and dispositions), organizational factors (structural-demographic characteristics and organizational behavior dimensions), student peer environment (campus racial and academic climates), and individual student experiences (curricular, classroom, and out-of-class experiences). Such reviews indicate the sheer complexity involved. Implementing an effective intervention is difficult due to complicated causes of attrition, embedded social inequalities, and expenses (Tinto 2006).

Interventions targeting non-academic factors are popular. For example, using a randomized controlled trial, Evans et al. (2020) implemented a case management intervention, where social workers helped students with financial assistance, course selection, finding childcare, and accessing social services. The case management intervention significantly improved persistence rates (Evans et al. 2020). First year programs that are separate from academic faculty are popular, but increasingly there is a focus on the classroom’s importance for retention (Tinto 2006). Tinto (1997) evaluated the impact of a college program designed for communal learning, which aims to engage students more than traditional programs. They found that students in the communal program participated in more academic activities, reported more positive views of the college and their role in it, and persisted at higher rates than the control (Tinto 1997).

Especially at the elementary and secondary level, many interventions to improve student outcomes do so by targeting ISRs. Roorda et al. (2011) performed a meta-analysis of 99 studies on teacher-student relationships and student engagement and achievement. Effect sizes were larger for engagement than for achievement, and the effect sizes varied widely based on the specific measurements being used (Roorda et al. 2011). They found stronger effects of teacher-student relationships in the upper secondary level, as compared to the lower elementary level (Roorda et al. 2011). At the college level, Tinto (2006) argues that the instructor is a key player in student retention. Creasey, Jarvis, and Gadke (2009) found that students who had connected, nonthreatening relationships with their instructors had more positive feelings about their own abilities and expectations for the course.

ISRs are built on a range of social behaviors and experiences, but similarity’s effect on ISRs is of particular interest for this study. This is a specific case of a wider field of study: how does similarity affect all types of relationships? In general, people are more likely to persist in building relationships with acquaintances that they share similarities with, including prejudices, behaviors, personality traits, attitudes, demographics, and activities (Bahns et al. 2017). In mentoring relationships, perceived deep similarity contributes to information sharing behavior and positive reception, which then is positively associated with mentee adjustment (Zheng et al. 2021).

Other studies consider similarity and student performance, but do not directly address ISRs. One study of Taiwanese high school students explored the link between shared education and life values and student performance (Lai 2015). While they found a significant, but small, positive relationship between life values and student performance on analytic tests, education values gave mixed results depending on the natures of the education value and the analytic test in question (Lai 2015). At the college level, Abrami and Mizener (1985)‘s study found a small correlation between students’ perceiving instructors’ attitudes to be similar to their own and the grades of those students, but the correlation became insignificant when instructor fixed effects were considered. In the college context, the link between similarity and grades is not well-established.

Beyond the wide range of research that addresses similarity, ISRs, and student outcomes, ‘Taking It to the

Next Level' was directly inspired by two previous works in middle and high schools. Gehlbach, Brinkworth, and Harris (2012)'s study of middle-school students found that students' perceived similarity to their teachers positively correlated with improvements in teacher-student relationship, and that changes in teacher-student relationships were associated with changes in student perceived self-efficacy, but not scholastic performance. Using a randomized controlled trial and surveys, Gehlbach et al. (2015) found that when high school students were presented with similarities they shared with their teachers, they subsequently perceived greater similarity with their teachers. When teachers were presented with similarities, they perceived better relationships with those students, and those students subsequently performed better academically. For students, however, awareness of similarities did not effectively improve their perception of the student-teacher relationship.

2 Methodology

Robinson, Scott, and Gottfried (2019) conducted a randomized controlled trial, with surveys, to assess the impact of awareness of instructor-student similarities on similarity perception, instructor-student relationship perception, grades, and reenrollment.

2.1 Participants

The study took place in the 2017 spring term at a large Californian University. The study included 148 instructors and their 3,355 students. The instructors participated in the study based on interest and a gift-card incentive, and their students were invited to participate unincentivized. Students were only enrolled in the study for one class, in the event that they were taking classes with multiple participating instructors. The initial sample of 148 instructors was a convenience sample: the study was advertised, interested instructors signed up and were enrolled if their course met the study requirements. 145 instructors consented and took the initial survey, resulting in an corresponding sample of 3,352 students. Once errors of administration, missing responses, and inadequate time spent on surveys were considered, the sample consisted of only 119 instructors and 2,273 students.

2.2 Treatment and control

Participating students were randomly assigned to either treatment or control. At the beginning of the term, all participating students and instructors were given "get to know you" surveys. Using those responses, for each student in the treatment group, seven commonalities were identified between student and instructor (for example, perhaps both student and instructor binge-watch TV to relieve stress, or appreciate loyalty as the most important friend quality), and both student and instructor were informed of these commonalities. They completed a few questions about the similarities and were reminded of them through the term to ensure they were internalized. Students in the control group were informed about similarities they shared with students in another part of the country, and instructors were told nothing about these students.

2.3 Procedures

At the beginning of the term, the instructors were surveyed, to collect potential similarities. Then, students were surveyed, to collect potential similarities. Immediately, students were either given the treatment or the placebo, and then given further survey questions that asked about similarity and ISR. Students were surveyed again at the end of the term, regarding similarity and ISR. Instructors were surveyed about similarity and ISR only at the end of the term.

2.4 Measures

Robinson, Scott, and Gottfried (2019) identify key measures. Full descriptions of all measures are available in Appendix A.

1. Immediately after the treatment or placebo, students answered six questions about their perceived similarity to the instructor, on scales of 1 to 5. These responses were averaged to create a *student perceived similarity scale.*

2. Immediately after the treatment or placebo, students answered seven questions about their anticipated instructor-student relationship, on scales of 1 to 5. These responses were averaged to create a *student anticipated ISR scale*.
3. At the end of the term, students answered the *student perceived similarity scale* questions again.
4. At the end of the term, students answered seven questions about their perception of the instructor-student relationship, on scales of 1 to 5. These responses were averaged to create a *student perceived ISR scale*.
5. At the end of the term, instructors answered only one question about *instructor perceived similarity* with the student, on a scale of 1 to 5.
6. At the end of the term, instructors answered seven questions about their perception of the instructor-student relationship, on scales of 1 to 5. These responses were averaged to create an *instructor ISR perception* scale.
7. Instructors were asked to report the student's *grade on their final exam*, paper, or project.
8. The *grade that the student received in the course* was pulled from internal records.
9. The student's course grade was *standardized* against other grades in the course.
10. The student's *status as of Fall term 2017*, either enrolled or not, was pulled from internal records.

2.4.1 Scale Reliability

A scale is externally reliable if it gives the same results across different testings (Coolican 2014). Test-retest reliability, where a scale is given to a group at different times, is a common way to assess this (Coolican 2014). In this data, there is a retest, but interactions between instructors and students may reasonably be expected to affect the second testing, rendering the data unuseful for assessing test-retest reliability.

A scale is internally reliable if it is consistent within itself (Coolican 2014). Cronbach's alpha was used to assess the reliability of the similarity and ISR scales. This statistic evaluates how much participants vary on individual items, compared to how they vary overall (Coolican 2014). Good reliability is indicated by an alpha between .75 and 1 (Coolican 2014). For the initial student perceived similarity scale, student anticipated ISR scale, end-of-term student perceived similarity scale, end-of-term student perceived ISR scale, and end-of-term instructor perceived ISR scale, $\alpha = 0.892$, 0.872, 0.91, 0.905, and 0.926, respectively, indicating that the scales are reliable.

2.4.2 Scale Validity

A scale is valid if it measures what it is supposed to measure (Coolican 2014). In this case, the scales should measure similarity perception and instructor-student relationship (ISR) perception. Robinson, Scott, and Gottfried (2019) used modified versions of the scales from Gehlbach et al. (2015), where they were used with ninth-grade American students and their teachers; the scales were originally were presented in Gehlbach, Brinkworth, and Harris (2012), and they were developed for use with sixth-, seventh-, and eight-grade American students and their teachers. Given the fact that these scales were developed for children, it is worth considering what the scales actually measure. To do this, we can look at the specific questions in the three distinct scales.

To assess similarity, students were asked about values, course goals, views on course content, general commonalities, personality, and overall similarity. By this standard, high similarity is when a student shares values, course goals, and views on course content, and the student and instructor have similar personalities, high commonality, and high general similarity. Questions ask about course-specific measures (content and goals), personal traits (values and personality), and non-specific traits (commonalities and general similarity), but do not ask about life experience, abilities, or approaches, amoungst other ways that two people can theoretically be similar.

To assess ISR, students were asked about enjoyment of learning, friendliness, encouragement, excitement, motivation, caring, and overall learning. By this standard, a high-quality ISR is one where a student enjoys learning from, is motivated by, and learns a great deal from a friendly, encouraging, caring instructor, who would be excited to see them in three years time. To assess ISR, instructors were asked about enjoyment of aiding learning, caring, frequency of encouragement, friendliness, excitement, motivation, and overall learning. By this standard, a high-quality ISR is one where an instructor enjoys helping and frequently encourages a friendly, caring student who is motivated and learns a great deal from this instructor, and the instructor would be excited to see the student in three years.

As Brinkworth et al. (2018) argue, the instructor-student relationship is conceptualized, and therefore measured, in a variety of ways. Other constructions of ISR scales may measure negative aspects, respect, liking, responsiveness, engagement, trust, honesty, humour, interest, approachability, clarity, confidence, fairness, patience, and consistency (Brinkworth et al. 2018). Robinson, Scott, and Gottfried (2019)'s ISR scale is measuring only one concept of an ISR, and that particular concept needs to be considered when analyzing the results of this experiment. Like most research about ISRs, Robinson, Scott, and Gottfried (2019)'s scale originates in elementary and secondary schools, and the concept of "good ISR" may not translate easily to a post-secondary context.

3 Data

This paper uses the R statistical language (R Core Team 2020) and `tidyverse` packages (Wickham et al. 2019). `here` was used to manage the RProject (Müller 2020). Stata files were handled using `readstata13` (Garbuszus and Jeworutzki 2018). The report was made using RMarkdown (Allaire et al. 2020) and `bookdown` (Xie 2020). Tables were made using `kableExtra` (Zhu 2020), `finalfit` (Harrison, Drake, and Ots 2020), and `stargazer` (Hlavac 2018). The correlation matrix was made using functions by Engineering (2018) using the `Hmisc` package (Harrell and Dupont 2020). `MASS` provided additional models (Venables and Ripley 2002). Plots used `cowplot` (Wilke 2020). `ltm` was used to check scale validity (Rizopoulos 2006). `tidymodels` (Kuhn and Wickham 2020) and `performance` (Lüdecke et al. 2021) were used in assessing model validity and performance. Finally, the diagram was made using `DiagrammeR` (Iannone 2020).

The data, in a .dta format, and Stata code for the Robinson, Scott, and Gottfried (2019) paper are available on the Inter-University Consortium for Political and Social Research (Robinson, Scott, and Gottfried 2020). The dataset contains 36,838 observations of 653 variables.

3.1 Demographics

Most key demographic traits come from the initial surveys given to students and instructors. For gender, they were asked to identify as male, female, or other, and given the option to enter text. For race/ethnicity, they were allowed to check any number of a range of options ('White/Caucasian', 'Black or African American', 'Hispanic American or Latino/a', 'Asian or Pacific Islander', 'American Indian or Alaskan', 'Middle Eastern', and 'Other Race'). They were also asked if they were the first generation in their family to attend college (no or yes). Finally, they were asked to enter their age in years. The student's year of study and cumulative CGPA as of the previous term and the class size came from the college's internal records.

Table 1 shows the student covariates for the treatment and control samples. There are no significant differences between the treatment and control groups, as indicated by p-values which are all above 0.4.

Table 2 shows the instructor covariates. Because instructors are counted multiple times, for each student in the treatment and control groups that is in their course, only the totals are shown. The same covariates, presented by treatment and control groups, are shown in Appendix B.

There are differences between the students and instructors. Notably, the student sample is 62.7% female, while the instructor sample is 78.2% female. The student sample is 19.8% White, while the instructor sample is 38.7% White. The student sample is 47% Hispanic, while the instructor sample is only 3.4%. Finally, the student sample is 9.8% multiracial, but the instructor sample is 52.1% multiracial. Some of these differences may be due to self-identification: while one person who is Hispanic may select only Hispanic, another may select, for example, Hispanic as well as White, and be classified differently as a result. For this reason, it is very difficult to draw conclusions. Additionally, students are 43.5% first-generation college students, while only 25.2% of instructors are.

In light of the different percentages in student and instructor racial self-identification, Figure 1 shows the percentages of each group that selected a given racial identity, as well as the percentage that selected

Table 1: Student covariates for treatment and control groups

Group	N	Missing		Control	Treatment	Total	p
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273	
Student gender	2255	18	Male	433 (38.4)	415 (36.8)	848 (37.6)	0.470
			Female	695 (61.6)	712 (63.2)	1407 (62.4)	
Student race/ethnicity	2273	0	American Indian or Alaskan	1 (0.1)		1 (0.0)	0.639
			Asian or Pacific Islander	126 (11.1)	124 (10.9)	250 (11.0)	
			Black or African American	51 (4.5)	43 (3.8)	94 (4.1)	
			Hispanic American or Latino/a	513 (45.1)	556 (48.9)	1069 (47.0)	
			Middle Eastern	54 (4.7)	55 (4.8)	109 (4.8)	
			Mixed race/ethnicity	113 (9.9)	111 (9.8)	224 (9.9)	
			Other Race	40 (3.5)	31 (2.7)	71 (3.1)	
			Unknown	2 (0.2)	3 (0.3)	5 (0.2)	
			White/Caucasian	237 (20.8)	213 (18.8)	450 (19.8)	
Student first-gen status	2270	3	No	634 (55.8)	648 (57.1)	1282 (56.5)	0.550
			Yes	502 (44.2)	486 (42.9)	988 (43.5)	
Year	2246	27	1	78 (6.9)	74 (6.6)	152 (6.8)	0.977
			2	178 (15.8)	180 (16.1)	358 (15.9)	
			3	191 (17.0)	182 (16.2)	373 (16.6)	
			4	665 (59.1)	668 (59.6)	1333 (59.3)	
			5	14 (1.2)	16 (1.4)	30 (1.3)	
Student age	2269	4	Mean (SD)	22.6 (5.4)	22.7 (5.3)	22.7 (5.4)	0.726
CPGA	2178	95	Mean (SD)	3.0 (0.6)	3.0 (0.7)	3.0 (0.6)	0.414

Table 2: Teacher covariates

	N	Missing		
				N = 119
Instructor gender	119	0	Male	26 (21.8)
			Female	93 (78.2)
Instructor race/ethnicity	119	0	Asian or Pacific Islander	4 (3.4)
			Black or African American	2 (1.7)
			Hispanic American or Latino/a	4 (3.4)
			Mixed race/ethnicity	62 (52.1)
			White/Caucasian	46 (38.7)
			Unknown	1 (0.8)
Instructor first-gen status	119	0	No	89 (74.8)
			Yes	30 (25.2)
Instructor age	118	1	Mean (SD)	44.5 (11.0)
Class size	118	1	Mean (SD)	36.4 (27.9)

more than one of these categories. While the tables display race/ethnicity as a single variable (it is either a single identity, mixed race/ethnicity, or unknown), this graph, and the subsequent analysis, deal with multiple racial/ethnic identities that are either selected or not. On the graph, we can see that more students identified as Hispanic, while more instructors identified as White and selected multiple options. Viewing only the table, where race is shown as a one-dimensional variable, does not show the nuance of multiple racial/ethnic identities. Still, these categories are limiting, and do not necessarily truly reflect the social or cultural identities of study participants.

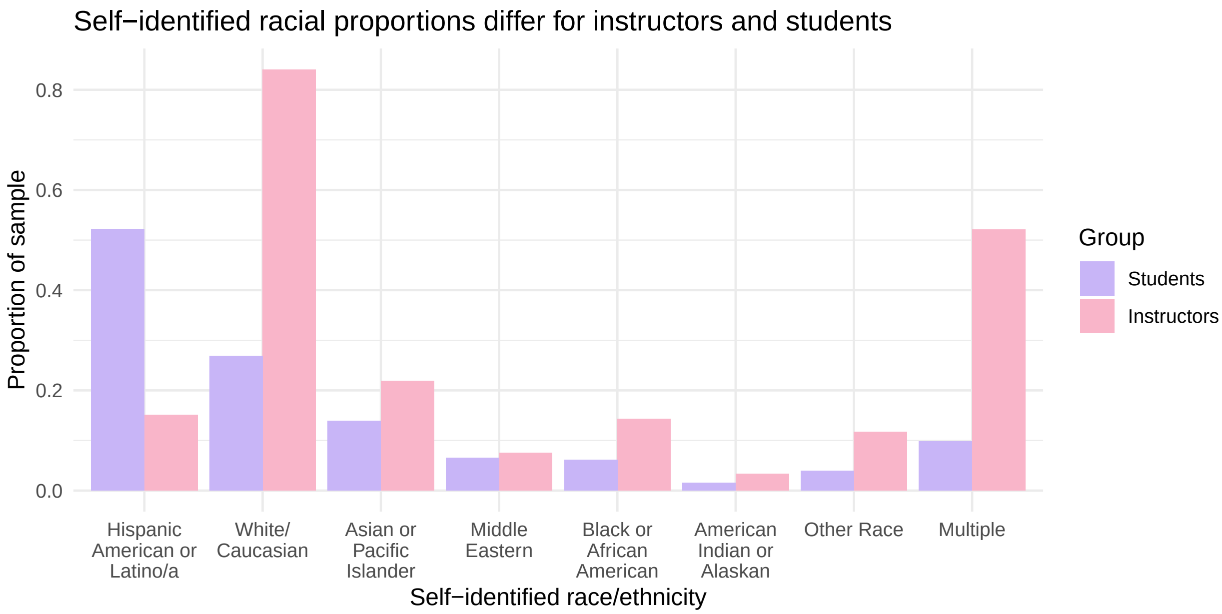


Figure 1: Bar chart illustrating racial identification proprtions of participants

Given the potential significance of class size for instructor-student relationships, Figure 2 shows the distribution of class sizes of instructors that participated in the study. Most class sizes are between 25 and 50 students, but there are some that are much larger.

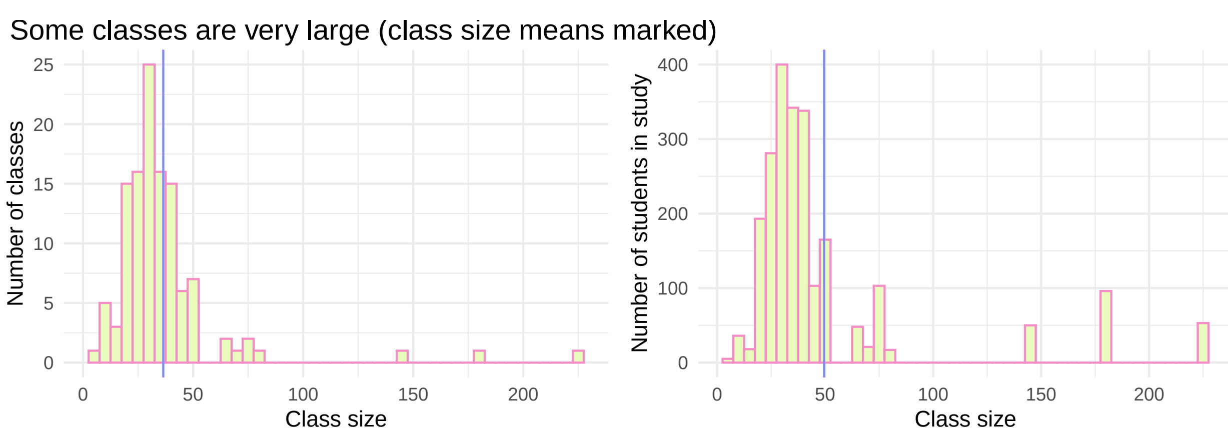


Figure 2: Histogram illustrating class sizes

Table 3: Outcomes of interest for treatment and control groups

	N	Missing		Control	Treatment	Total	p
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273	
Initial student perception							
Similarity	2272	1	Mean (SD)	3.4 (0.7)	3.6 (0.7)	3.5 (0.7)	<0.001
ISR	2273	0	Mean (SD)	4.1 (0.6)	4.1 (0.6)	4.1 (0.6)	0.997
End-of-term student perception							
Similarity	1831	442	Mean (SD)	3.5 (0.8)	3.6 (0.8)	3.6 (0.8)	0.004
ISR	1831	442	Mean (SD)	4.1 (0.7)	4.1 (0.7)	4.1 (0.7)	0.888
End-of-term instructor perception							
Similarity	2225	48	1	133 (11.9)	132 (11.9)	265 (11.9)	1.000
			2	344 (30.8)	343 (31.0)	687 (30.9)	
			3	402 (36.0)	400 (36.1)	802 (36.0)	
			4	171 (15.3)	166 (15.0)	337 (15.1)	
			5	67 (6.0)	67 (6.0)	134 (6.0)	
ISR	2240	33	Mean (SD)	3.5 (0.9)	3.5 (0.9)	3.5 (0.9)	0.639
Student outcomes							
Course grade	2147	126	Mean (SD)	3.1 (1.0)	3.1 (1.0)	3.1 (1.0)	0.702
Stand. grade	2120	153	Mean (SD)	0.1 (0.9)	0.1 (0.9)	0.1 (0.9)	0.517
Final grade	2109	164	Mean (SD)	3.0 (1.1)	3.0 (1.1)	3.0 (1.1)	0.525
Peristence	1786	487	No	59 (6.7)	68 (7.5)	127 (7.1)	0.553
			Yes	823 (93.3)	836 (92.5)	1659 (92.9)	

3.2 Outcomes of interest[1]

Table 3 shows summary statistics for key variables identified by Robinson, Scott, and Gottfried (2019), including missing values. There do not appear to be significant differences between the treatment and control groups, other than in student similarity perception ($p < 0.001$ for initial perception, and $p = 0.004$ for end-of-term perception). Even in these two measures, the difference appears to be small (mean scores of 3.4 versus 3.6 for intial perception and 3.5 versus 3.6 for end-of-term perception).

Figure 3 shows key continuous outcome distributions for both treatment and control. Corresponding to Table 3, the only visible differences in distributions are in the initial student similarity perception and end-of-term student similarity perception.

Interestingly, there is a large difference between how students rate perceived similarity and how instructors rate perceived similarity. Students, on average, rate end of term similarity at 3.6, while instructors give an average rating of 2.7. Students give an average end-of-term ISR rating of 4.1, while instructors give an average of only 3.5. This could be due to more positive student perceptions, or due to students giving higher ratings for equivalent perceptions. Both groups rate ISR more highly than similarity, but it is difficult to compare the weight of these scores as the scales are not necessarily measuring in the same units.

Replicating Robinson, Scott, and Gottfried (2019)'s study, Table 4 displays a correlation matrix for the measures that Robinson, Scott, and Gottfried (2019) identified as key. The very strong correlations (course grade and the standardized version, course grade and final grade, and course grade and objectively graded final grade) and some of the strong correlations (standardized course grade with final grade and objectively grade final grade) are entirely expected.

Other strong correlations are more interesting: initial student similarity and instructor-student relationship perceptions, end-of-term student similarity and instructor-student relationship perceptions, and end-of-term instructor similarity and instructor-student relationship perceptions. At a given moment in time, a participant's assessment of similarity and ISR are highly related. This could indicate that the two perception scales may not be measuring psychologically distinct concepts, or that feelings of similarity and good relationships are strongly associated in the classroom.

[1] Interactive graphs and tables showing outcomes of interest are available here.

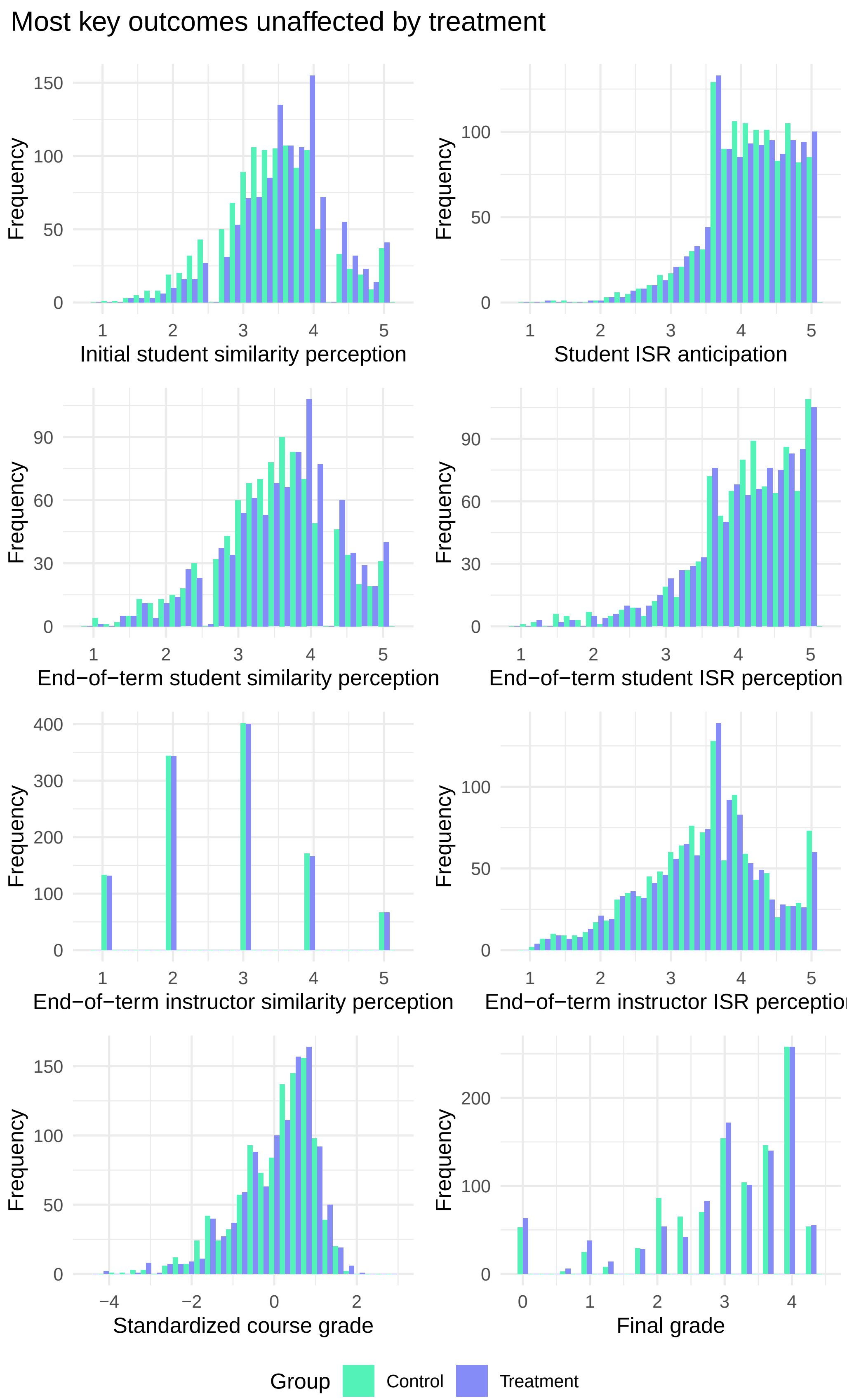


Figure 3: Histograms illustrating continous and ordinal key outcomes

Table 4: Correlation matrix for continuous and ordinal key variables and outcomes

	N	Missing	Mean (SD)	Initial student Similarity	Initial student ISR	End-of-term student Similarity	End-of-term student ISR	End-of-term instructor Similarity	End-of-term instructor ISR	Grades Course	Grades Stand. course
Initial student perception											
Similarity	2272	1	3.5 (0.7)	-							
ISR	2273	0	4.1 (0.6)	0.67***	-						
End-of-term student perception											
Similarity	1831	442	3.6 (0.8)	0.58***	0.5***	-					
ISR	1831	442	4.1 (0.7)	0.43***	0.57***	0.79***	-				
End-of-term instructor perception											
Similarity	2225	48	2.7 (1.0)	0.15***	0.16***	0.25***	0.24***	-			
ISR	2240	33	3.5 (0.9)	0.19***	0.21***	0.29***	0.33***	0.68***	-		
Grades											
Course	2147	126	3.1 (1.0)	0.07**	0.1***	0.2***	0.23***	0.38***	0.44***	-	
Stand. course	2120	153	0.1 (0.9)	0.02	0.01	0.12***	0.11***	0.29***	0.35***	0.81***	-
Final	2109	164	3.0 (1.1)	0.06**	0.1***	0.17***	0.18***	0.38***	0.43***	0.81***	0.62***
Obj. Final	538	1735	2.8 (1.2)	0.05	0.07	0.17***	0.15**	0.43***	0.52***	0.83***	0.64***

Moderate correlations exist between initial and end-of-term student similarity perceptions as well as initial and end-of-term student ISR perceptions, indicating that student perceptions change somewhat but not entirely through the term. There are some moderate correlations between instructor perceptions of similarity or ISR and course grade, final grade, and objectively graded final grade, perhaps indicating that instructor opinions of students are linked to student performance in some way.

3.3 Missing data

There are no significant patterns in missing data. Appendix C shows the distributions of key variables relative to missing data in other key variables.

3.4 Attrition

140 eligible instructors completed the first survey, and 2,808 of their students completed the first survey. 21 of those instructors did not complete the second survey. 2,808 students completed the first survey (1,396 in the treatment group, 1,412 in the control group), but 755 of those students did not complete the second survey (364 in the treatment group, 391 in the control group). The attrition in the treatment and control groups was comparable (26.1% in the treatment group, 27.7% in the control group).

3.5 Data selection

From the 36,838 observations, or potential units of study (corresponding to all undergraduate student records of the university), only 2,273 were used in analysis. Units of study were excluded because the instructor did not participate in the study (33,483); the instructor or course wasn't actually eligible (96); the instructor administered the wrong survey (31); the student did not consent (116); the student did not complete the first survey (304); the instructor did not complete the second survey (107); or one participant did not spend at more than a second reading the consent page and more than ten seconds reading each page with five questions or more (428). The time limits are somewhat arbitrary, and different choices could have different analysis results.

4 Models

4.1 Replication models

$treatment_i$ is the indicator that treatment was given.

X_{1i} is a vector of student-level covariates (student ISR anticipation, gender, CGPA, year of study). Different covariates are used to model different outcomes. The original study used gender and CGPA are used as controls for grade based outcomes only. Robinson, Scott, and Gottfried (2019) state this is because females earn higher grades generally. This paper uses them as controls for the majority of outcomes. ISR anticipation is used as a control only for end-of-term student ISR perception.

X_{2j} is a vector of instructor-level covariates (class size). Different covariates are used to model different outcomes. The original study consider instructor fixed effects instead. Because each participating instructor taught one class in the study, the class sizes correspond directly to the instructor IDs, so it was inadvisable to use both. This paper selected class size because of the larger potential for transferable insights.

ϵ_{ij} is a standard error. As this paper does not use clustered standard errors like Robinson, Scott, and Gottfried (2019), there are slight differences in standard error and p-values when compared with the original paper. This does not affect the regression coefficients.

In the resulting models, β_0 and a_k are constant terms. β_1 is the coefficient that indicates treatment effect. Γ_1 and Γ_2 are coefficients for the covariates.

4.1.1 Linear models

Equation (1) is Robinson, Scott, and Gottfried (2019)'s linear model, used for continuous outcomes. This include complete scale outcomes: initial student similarity perception, end-of-term student similarity perception, end-of-term student ISR perception, and end-of-term instructor ISR perception. Because these scales are created using 6-7 questions answered on a scale of 1 to 5, they have a total of 25 or 30 possible values, they can be treated as continuous variables and modeled using linear functions. It also includes grade-based outcomes: course grade and objectively graded exam grade, both of which are shown on a 4.0 GPA scale.

$$Outcome_{ij} = \beta_0 + \beta_1 treatment_i + \Gamma_1 X_{1i} + \Gamma_2 X_{2j} + \epsilon_{ij} \tag{1}$$

4.1.2 Ordinal logistic models

Equation (2) is Robinson, Scott, and Gottfried (2019)'s ordinal logistic model, used for the ordinal outcome, which is end-of-term instructor similarity perception. Because the instructors were only asked to answer one five-point scale question about similarity, this outcome cannot be treated the same way as the full-scale outcomes.

$$prob(outcome_{ij}) = a_k + \beta_1 treatment_i + \Gamma_1 X_{1i} + \Gamma_2 X_{2j} + \epsilon_{ij} > k \tag{2}$$

4.1.3 Logistic models

Equation (3) is Robinson, Scott, and Gottfried (2019)'s logistic model, used for the binary outcome, which is enrollment in Fall term 2017. Students are either not enrolled (0), enrolled (1), or they have graduated and are excluded from the model.

$$prob(outcome_{ij}) = a_k + \beta_1 treatment_i + \Gamma_1 X_{1i} + \Gamma_2 X_{2j} + \epsilon_{ij} > k \tag{3}$$

4.2 Additional models

When creating additional models, I focused on three key outcomes: standardized course grade, end-of-term student perceived ISR scale, and end-of-term student perceived similarity scale. I selected these because interventions like Robinson, Scott, and Gottfried (2019)'s are aimed at improving student outcomes (like course performance), and it is much more likely to detect an impact in the immediate course outcomes than outcomes from subsequent terms; and because previous studies show more detectable impact on student perception than instructor perception. Equation (4) shows the linear model used. X_{3jk} represents student-instructor commonality covariates (matching racial self-ID, matching gender self-ID, age difference).

$$Outcome_{ij} = \beta_0 + \beta_1 treatment_i + X_{1i}\Gamma_1 + X_{2j}\Gamma_2 + X_{3jk}\Gamma_3 + \epsilon_{ij} \tag{4}$$

5 Results

All results are reported using the treatment coefficient, β_1, and covariate coefficients, which are represented with Γ. 95% confidence intervals for those coefficients are shown in brackets below the coefficients. Significance levels are indicated with stars. Initial perceptions are indicated with 1, and end-of-term perceptions with 2. ISR refers to instructor-student relationship perception, and sim. refers to similarity perception. Final grade refers to the grade given for an objectively graded final exam in the course, and Std. course grade refers to a standardized final course grade.

5.1 Replication results

In Table 5, Model 1 shows that the treatment had a statistically significant effect on initial student similarity perception ($\beta_1 = 0.194$, 95% CI [0.136, 0.253]). Similarly, Model 2 that the treatment had a statistically significant effect on student end-of-term similarity perception ($\beta_1 = 0.11$, 95% CI [0.038, 0.183]), though this effect size is smaller. Both coefficients are minimal on a scale of 1 to 5. The adjusted R-squared values, which indicate the proportion of variation in the outcome that can be explained by the treatment and covariates, are low (0.032 and 0.022). These models are helpful for analyzing treatment effect, but they do not aim to be and are not useful predictors of student similarity perception, not even within the sample.

Model 3 shows no evidence that treatment impacted student end-of-term ISR perception. However, it does show a significant association between student ISR anticipation and student end-of-term ISR perception ($\Gamma = 0.671$, 95% CI [0.624, 0.717]). For every 1 point increase in ISR anticipation on a 1 to 5 scale, the expected increase in end-of-term ISR perception is 0.671. Models 4, 5, 6, 7, and 8 show no significant treatment effect on end-of-term instructor ISR perception, course grade, objectively graded exam grade, end-of-term instructor similarity perception, or enrollment in the subsequent term.

All models in Table 5 use gender, prior GPA, and year of study as student-level covariates. Instructor-level covariates are limited to the class size. This is based on the assumption that student traits will be more likely to affect student experiences, and therefore student outcomes. Race/ethnicity is not included because that covariate is further explored in separate models.

Model performance and validation for the two models with significant treatment effects, Models 1 and 2 from Table 5, can be found in Appendix D.

Because some students are more commonly targeted by persistence interventions, the subset of Hispanic and/or Black college student participants is considered separately from the rest of the student population. While the effect of treatment on student similarity perception is larger than for the general student population ($\beta_1 = 0.199$, 95% CI [0.124, 0.274] for initial perception; $\beta_1 = 0.116$, 95% CI [0.021, 0.211] for end-of-term perception), there is still no effect on ISR perception, instructor perception, or student outcomes. Full results are in Appendix E.

Similarly, the subset of first-generation college student participants is considered. The effect of treatment on student similarity perception is larger than for the general student population and the Black and/or Hispanic student participants ($\beta_1 = 0.265$, 95% CI [0.177, 0.353] for initial perception; $\beta_1 = 0.15$, 95% CI [0.039, 0.26] for end-of-term perception), but there is still no effect on ISR perception, instructor perception, or student outcomes. Full results are in Appendix F.

5.1.1 Exploratory replication results

Table 6 disregards the treatment condition, instead considering the association between student identity and similarity/ISR. While we cannot make causal inferences, this is an exploratory analysis that could shape future intervention studies. If a student identifies as first generation, can we anticipate a different student or instructor ISR perception? In this case, there is no significant association.

Table 5: Replication model results

	Dependent variable:							
	Student sim. 1	Student sim. 2	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Instructor sim. 2	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*ordered logistic*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	$.192^{***}$	$.108^{***}$	$-.001$	$-.014$	$.027$	$-.021$	$-.021$	$.117$
	(.134,.251)	(.035,.181)	(-.058,.056)	(-.086,.057)	(-.049,.104)	(-.215,.174)	(-.176,.133)	(-.302,.535)
Student ISR 1			$.674^{***}$					
			(.627,.721)					
Class size	$.0004$	$.001^{*}$	$-.00001$	$.003^{***}$	$.0003$	$.0001$	$.009^{***}$	$.004$
	(-.0003,.001)	(-.0001,.002)	(-.001,.001)	(.002,.003)	(-.001,.001)	(-.002,.002)	(.007,.011)	(-.001,.009)
Student female	$.093^{***}$	$.063$	$.002$	$.030$	$.074^{*}$	$-.153$	$.289^{***}$	$-.204$
	(.032,.154)	(-.014,.140)	(-.059,.062)	(-.045,.105)	(-.006,.155)	(-.364,.059)	(.125,.452)	(-.645,.237)
CGPA	$.029$	$.076^{***}$	$.030$	$.244^{***}$	$.624^{***}$	$.482^{***}$	$.543^{***}$	$.895^{***}$
	(-.017,.075)	(.018,.134)	(-.015,.075)	(.188,.301)	(.564,.684)	(.346,.619)	(.416,.671)	(.640,1.150)
Year of study	$.065^{***}$	$.087^{***}$	$.057^{***}$	$.062^{***}$	$.142^{***}$	$.219^{***}$	$.116^{***}$	$.055$
	(.035,.096)	(.048,.126)	(.027,.088)	(.025,.100)	(.103,.182)	(.120,.318)	(.034,.198)	(-.147,.257)
Constant	3.030^{***}	2.889^{***}	1.049^{***}	2.429^{***}	$.673^{***}$	$.814^{***}$		$.021$
	(2.861,3.199)	(2.671,3.107)	(.802,1.296)	(2.221,2.638)	(.454,.892)	(.318,1.310)		(-.865,.906)
Observations	2,160	1,747	1,747	2,134	2,063	514	2,119	1,687
R^2	.034	.025	.327	.057	.210	.150		
Adjusted R^2	.032	.022	.324	.054	.208	.141		
Akaike Inf. Crit.								709.967

*p<0.1; **p<0.05; ***p<0.01

Table 6: Replication model results: first-generation status and ISR

	Dependent variable:		
	Student ISR 1	Student ISR 2	Instructor ISR 2
	(1)	(2)	(3)
Student first gen.	.018	.003	$-$.047
	(-.036,.072)	(-.067,.074)	(-.116,.022)
Student female	.110***	.066*	.026
	(.055,.166)	(-.007,.138)	(-.044,.097)
CGPA	$-$.031	$-$.067**	.003
	(-.077,.014)	(-.126,-.008)	(-.055,.061)
Course grade	.056***	.197***	.390***
	(.026,.086)	(.153,.241)	(.352,.428)
Year of study	.048***	.084***	.008
	(.021,.076)	(.047,.121)	(-.027,.044)
Class size	.0004	.0003	.002***
	(-.0002,.001)	(-.001,.001)	(.001,.003)
Constant	3.797***	3.348***	2.152***
	(3.642,3.952)	(3.133,3.564)	(1.952,2.353)
Observations	2,061	1,674	2,034
R^2	.025	.067	.214
Adjusted R^2	.022	.063	.211

*p<0.1; **p<0.05; ***p<0.01

Table 7, however, shows that there is a significant negative association between student identification as Hispanic and/or Black and end-of-term instructor ISR perception (Γ = -0.096, 95% CI [-0.166, -0.026]).

Table 7: Replication model results: Hispanic and/or Black students and ISR

	Dependent variable:		
	Student ISR 1	Student ISR 2	Instructor ISR 2
	(1)	(2)	(3)
Student Hispanic and/or Black	$-$.026	.019	$-$.094***
	(-.081,.029)	(-.054,.092)	(-.164,-.023)
Student female	.115***	.065*	.035
	(.059,.170)	(-.008,.138)	(-.036,.106)
CGPA	$-$.035	$-$.066**	$-$.003
	(-.081,.010)	(-.125,-.006)	(-.061,.055)
Course grade	.055***	.199***	.388***
	(.025,.084)	(.155,.243)	(.350,.427)
Year of study	.046***	.085***	.004
	(.018,.073)	(.048,.122)	(-.032,.039)
Class size	.0004	.0002	.002***
	(-.0002,.001)	(-.001,.001)	(.001,.003)
Constant	3.840***	3.328***	2.220***
	(3.679,4.001)	(3.104,3.553)	(2.011,2.428)
Observations	2,063	1,676	2,036
R^2	.025	.067	.216
Adjusted R^2	.022	.064	.214

*p<0.1; **p<0.05; ***p<0.01

Table 8 considers the association between similarity perception and ISR perception and the relationship between ISR perception and student outcomes. Again, because only the treatment was randomized, we cannot make causal inferences. Course size is used as a covariate for ISR perception because it directly shapes how students and instructors relate to one another. Student gender and prior GPA are used as covariates for student outcomes for the reasons previously outlined from Robinson, Scott, and Gottfried (2019).

In Models 1 and 2, we can see that student and instructor similarity perception are both significantly associated with student and instructor ISR perception. Unsurprisingly, student similarity perception is more strongly associated with student ISR perception ($\Gamma = 0.733$, 95% CI [0.706, 0.76]), and instructor similarity perception is more strongly associated with instructor ISR perception ($\Gamma = 0.512$, 95% CI [0.483, 0.541]).

Models 3 and 4 show that student ISR perception is significantly associated course grade, while instructor ISR perception is significantly associated with course grade and final grade. Instructor perception is more strongly associated ($\Gamma = 0.34$, 95% CI [0.295, 0.385] for course grade; $\Gamma = 0.643$, 95% CI [0.642, 0.759] for final grade) than student perception ($\Gamma = 0.115$, 95% CI [0.064, 0.164] for course grade).

Model 5 indicates that ISR perception is not significantly associated with persistence.

Table 8: Replication model results: Similarity, ISR, and student outcomes

	Dependent variable:				
	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*logistic*
	(1)	(2)	(3)	(4)	(5)
Treatment	$-.086^{***}$	$-.037$	.017	$-.078$	.249
	(-.128,-.045)	(-.094,.019)	(-.053,.087)	(-.263,.108)	(-.290,.788)
Student sim. 2	$.733^{***}$	$.146^{***}$			
	(.706,.760)	(.109,.182)			
Instructor sim. 2	$.043^{***}$	$.512^{***}$			
	(.022,.065)	(.483,.541)			
Class size	$-.001^{**}$	$-.001^{**}$			
	(-.001,-.0001)	(-.001,-.00002)			
Student ISR 2			$.114^{***}$	.072	.044
			(.064,.164)	(-.050,.195)	(-.312,.401)
Instructor ISR 2			$.340^{***}$	$.642^{***}$	$-.033$
			(.295,.385)	(.525,.759)	(-.367,.301)
Student female			.050	$-.051$	.048
			(-.022,.122)	(-.252,.150)	(-.503,.599)
CGPA			$.472^{***}$	$.280^{***}$	$.710^{***}$
			(.417,.527)	(.153,.407)	(.403,1.016)
Constant	1.476^{***}	1.674^{***}	.053	$-.358$	.874
	(1.373,1.579)	(1.535,1.814)	(-.209,.315)	(-1.008,.292)	(-.857,2.605)
Observations	1,809	1,809	1,665	394	1,343
R^2	.632	.460	.304	.304	
Adjusted R^2	.631	.459	.302	.295	
Akaike Inf. Crit.					465.174

*p<0.1; **p<0.05; ***p<0.01

5.2 Additional results

Table 9 shows the results of models that consider matching student-instructor traits. As with the replication results, there is a measurable treatment effect for initial student similarity perception ($\beta_1 = 0.108$, 95% CI [0.036, 0.179]). However, the coefficient is even lower than in the replication models. There is still no treatment effect for student end-of-term ISR perception or standardized course grade. Because the matching

ID covariates were used, student covariates like gender and race were not included. Prior GPA and class size were still included, because they do not have a "matching" trait.

Table 9: Additional model results

	Dependent variable:		
	Student sim. 2	Student ISR 2	Std. course grade
	(1)	(2)	(3)
Treatment	.102***	−.009	.030
	(.030,.174)	(-.077,.059)	(-.047,.108)
CGPA			.651***
			(.592,.711)
Class size	.0004	−.0001	−.0003
	(-.0005,.001)	(-.001,.001)	(-.001,.001)
Matching racial self-ID	−.022	−.110*	.166***
	(-.140,.096)	(-.221,.001)	(.040,.293)
Matching gender self-ID	.071*	.053	.108***
	(-.004,.145)	(-.017,.124)	(.028,.188)
Age difference	−.007***	−.006***	−.008***
	(-.010,-.004)	(-.009,-.004)	(-.011,-.005)
Number of similarities	.011*	.011*	−.005
	(-.002,.025)	(-.001,.024)	(-.019,.009)
Constant	3.436***	4.094***	1.290***
	(3.228,3.643)	(3.898,4.289)	(1.005,1.576)
Observations	1,804	1,804	2,051
R^2	.021	.017	.202
Adjusted R^2	.018	.013	.199

*p<0.1; **p<0.05; ***p<0.01

6 Discussion

The results of this replication study align with Robinson, Scott, and Gottfried (2019)'s findings. When undergraduate students and their instructors learn about similarities they share, there is a small improvement in student similarity perception. However, the treatment has no significant impact on student ISR perception, instructor perception, or student outcomes.

To understand what was and wasn't impacted by the treatment, we can consider the desired causal sequence in this study, shown in Figure 4. The intervention refers to informing students and instructors about shared traits. It does not target actual instructor-student similarity – it simply aims to heighten awareness of any existing similarities. Similarity perception is measured by the two perceived similarity scales. ISR, or instructor-student relationship, cannot be directly measured – only perception can be measured, through the perceived ISR scales. The student outcomes, grades and persistence, are more easily measured. The diagram shows the complexity of succeeding with this type of intervention: the treatment must affect perception, which must then affect relationships, which must then affect student outcomes.

This study did successfully demonstrate relationship 1: the treatment caused a statistically significant increase in student similarity perception at the beginning and end of the term. Relationship 2 failed to materialize, as there was no significant difference in instructor similarity perception in the treatment group. To understand why the treatment effect was small for students and insignificant for instructors, we can consider the specifics of the treatment. What similarities were students and instructors informed of? Appendix G contains a complete list. Students and instructors were presented with generalities such as "An instructor's most important characteristic? You both say being knowledgeable" or "You are cut from the same cloth–you both described yourself as thoughtful!" Such facts may not be unusual enough to be memorable, reducing

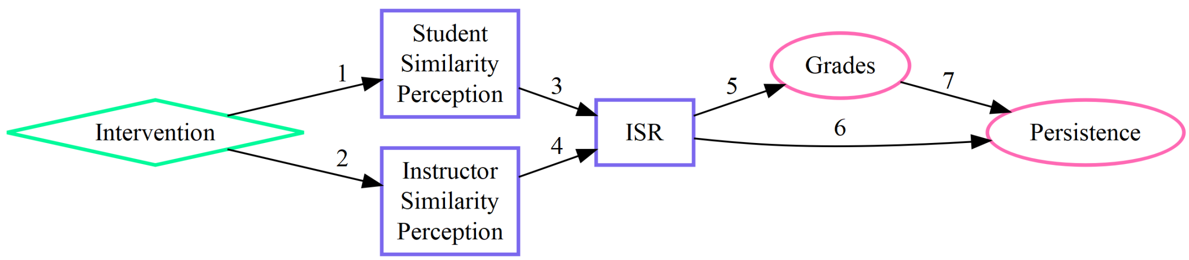


Figure 4: Diagram of desired causal sequence

the treatment's impact. It is also possible that similarities declared by an anonymous third party are simply not as impactful as similarities that are personally observed and experienced.

The exploratory modeling in Table 8 indicates that relationships 3 and 4 may exist: increases in similarity perception were associated with statistically significant increases in ISR perception. However, this relationship was not the focus of the study, and we cannot draw any firm conclusions. We also cannot determine causality: it is possible that an improved instructor-student relationship results in an increase in perceived similarity, not vice-versa. Additionally, much of the research about similarity and ISRs comes from elementary and secondary school contexts. The exact relationship between perceived similarity and ISRs could be different at the college level.

The exploratory modeling in Table 8 also indicates that ISR, or at least ISR perception, is related to student outcomes, as in relationships 5 and 6 in Figure 4. Again, we cannot conclude anything about causality. It is possible that students engage in behaviors that both earn higher grades and build better relationships with their instructors. Because the quality of the ISR cannot be measured directly, it is also possible that instructor-student relationships can improve student outcomes, but the particular perceptions that are measured in this case are not the relevant factors, or perceived ISR is less important that the real behaviors within an ISR.

The environment of Robinson, Scott, and Gottfried (2019)'s experiment is also relevant. They aimed to build ISRs in a context where a student's average class size is 49.5, students typically take five classes at a time, and instructors teach multiple classes, sometimes at multiple campuses. The intervention is appealingly simple, quick, and inexpensive, but such a small change is likely insignificant in the face of a wider campus culture that makes connections challenging.

Further, the study indicates that the instructor and student populations are demographically different. Compared to the group of instructors, the student body had 3.5 times as many individuals who identified as Hispanic. Compared to students, the instructor population had 3.1 times as many individuals who identified as White. This is only a hint of the differences that may exist between the student and instructor populations. As most students will not go on to be instructors, we can expect, at the very least, that many students will not share life goals and trajectories with their instructors. If students feel fundamentally different from their instructors, based on life experience or socio-economic status, an intervention that provides superficial similarities is insufficient to bridge that gap.

If colleges are considering investing in interventions that try to enhance similarity perception, we need to ask what it means to perceive similarity. Do some similarities matter more than others in ISRs? As indicated by the results in Table @ref(tab:modeltablerel2, some marginalized student identities may be associated with measurably lower ISR perceptions. If demographics traits, which are already part of an inequitable campus environment, are important for perceived similarity, should positive ISRs be contingent on similarity at all? There are other ways to cultivate ISRs that may make more sense and offer more equitable outcomes.

6.1 Limitations

The study's internal validity comes from the successfully randomly-assigned treatment and control groups, but there is one concern. The instructors who participated were part of both the treatment and control group. That is, they received similarities for some students but not others. If more significant effect sizes, or larger effect sizes, had been found, this potential lack of separation between the treatment and control groups would be a problem, because the instructor's attention could have simply been shifted from students in the control group to students in the treatment group, potentially exaggerating the effect. However, as little treatment effect was found, this limitation is less concerning.

The external validity is more questionable. The study participants were chosen in two main ways: the instructor was interested in the study and completed all the steps successfully, and the student agreed to participate and followed through. The sample of instructors that would be interested in the study would not necessarily be comparable to the overall population: they could be instructors who care more about supporting the work of other academics, instructors for whom a $150 giftcard was a significant draw, or instructors who are particularly interested in the role of ISRs.

Any measurable effects could also be specific to this particular college campus. Interventions may not be transferable due to different demographics, cultures, and institutional structures.

As with all survey studies, there is a risk that self-reported information is inaccurate. In particular, the surveys used by Robinson, Scott, and Gottfried (2019) had many Likert scale questions about very similar topics. Survey fatigue may have reduced the meaningfulness of the survey answers, especially those later in the survey. While responses that were given overly quickly were removed from the analysis, the fact that 428 cases had to be removed for this reason indicates a common problem with survey data: we cannot know if the participants actually considered their answers.

6.2 Future directions

Given the lack of significant findings, future studies could take a wide variety of approaches. First, more intensive, but similar, interventions to increase perceived similarity could be tested. These interventions could include more specific or important similarities between students and instructors. Second, different methods of increasing perceived similarity, perhaps based in actual instructor-student interaction, could be tested. Such interventions are less desirable from the college's point of view, because they would be more costly and demanding to implement. Third, interventions that aim to improve ISR without increasing perceived similarity could be considered. Similarity is only one of many theorized methods for improved ISRs.

A Appendix A

Complete list of key measures

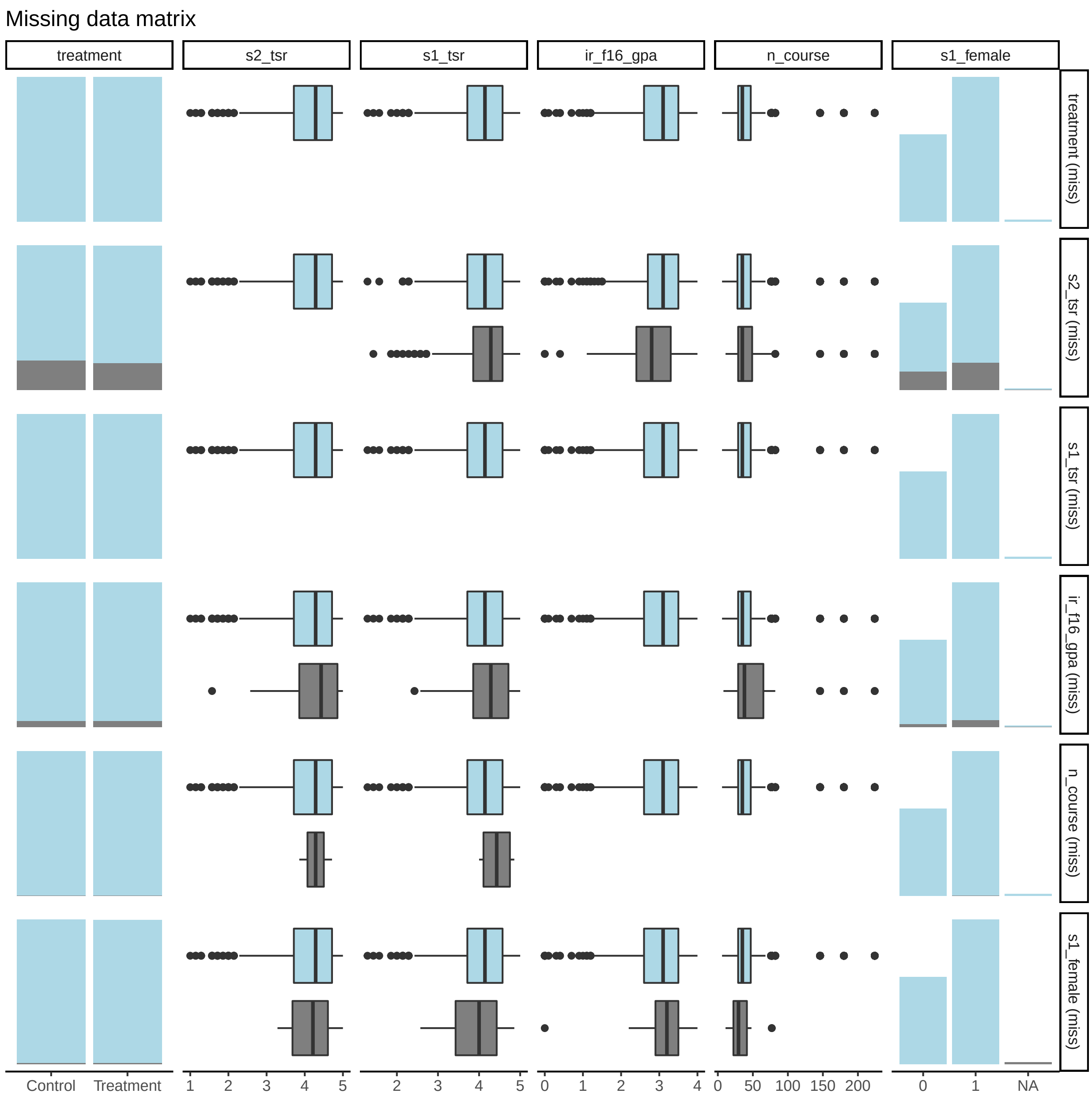
- Initial student survey
  1. Student perceived similarity scale
     - Overall, how similar to your instructor's values do you think your values are?
     - How similar are your goals for the course and your instructor's goals?
     - In general, how similar do you think your views about the course content and your instructor's are?
     - How much do you think you have in common with your instructor?
     - How similar do you think your personality is compared to your instructor's?
     - Overall, how similar do you think you and your instructor are?
  2. Student anticipated ISR scale
     - How much do you think you will enjoy learning from this instructor?
     - How friendly do you think this instructor will be towards you?
     - How encouraging do you think this instructor will be towards you?
     - If you came back to visit this instructor three years from now, how excited do you think they would be?
     - How motivating do you think you will find this instructor's class?
     - How caring do you think this instructor will be towards you?
     - Overall, how much do you think you will learn from this instructor?
- End of term student survey
  3. Student perceived similarity scale
  4. Student perceived ISR scale
     - How much do you enjoy learning from this professor?
     - How friendly do you think this professor is towards you?
     - If you came back to visit this professor three years from now, how excited do you think they would be?
     - How motivating do you find this professor's class?
     - How caring do you think this professor is towards you?
     - How encouraging do you think this professor is towards you?
     - Overall, how much do you think you have learned from this professor?
- End of term instructor survey
  5. Instructor similarity perception
     - Overall, how similar do you think you and STUDENTNAME are?
  6. Instructor perceived ISR scale
     - How much did you enjoy helping STUDENTNAME learn?
     - How caring was STUDENTNAME towards you?
     - How often did you say something encouraging to STUDENTNAME?
     - How friendly was STUDENTNAME towards you?
     - If this student came back to visit you three years from now, how excited would you be?
     - How motivating did STUDENTNAME find the activities that you plan for class?
     - Overall, how much did STUDENTNAME learn from you?
  7. Final grade: Instructors were asked to report the student's grade on their final exam, paper, or project.
- University internal records:
  9. Course grade: The final grade that the student received in the course.
  10. Standardized course grade: The student's final grade, standardized against other grades in the course.
  11. Persistence: The student's status as of Fall term 2017: not enrolled or enrolled.

B Appendix B

Table 10: Teacher covariates for treatment and control groups

Group	N	Missing		Control	Treatment	Total	p
				N(%) = 1137 (50.0)	N(%) = 1136 (50.0)	N(%) = 2273	
Instructor gender	2273	0	Male	230 (20.2)	229 (20.2)	459 (20.2)	1.000
			Female	907 (79.8)	907 (79.8)	1814 (79.8)	
Instructor race/ethnicity	2273	0	Asian or Pacific Islander	37 (3.3)	35 (3.1)	72 (3.2)	0.986
			Black or African American	7 (0.6)	5 (0.4)	12 (0.5)	
			Hispanic American or Latino/a	35 (3.1)	34 (3.0)	69 (3.0)	
			Mixed race/ethnicity	615 (54.1)	621 (54.7)	1236 (54.4)	
			Unknown	24 (2.1)	21 (1.8)	45 (2.0)	
			White/Caucasian	419 (36.9)	420 (37.0)	839 (36.9)	
Instructor first-gen status	2273	0	No	891 (78.4)	890 (78.3)	1781 (78.4)	1.000
			Yes	246 (21.6)	246 (21.7)	492 (21.6)	
Instructor age	2263	10	Mean (SD)	43.4 (10.5)	43.4 (10.6)	43.4 (10.5)	0.958
Class size	2269	4	Mean (SD)	49.3 (44.3)	49.7 (44.9)	49.5 (44.6)	0.812

C Appendix C


Missing data matrix
treatment
s2_tsr
s1_tsr
ir_f16_gpa
n_course
s1_female
treatment (miss)
s2_tsr (miss)
s1_tsr (miss)
ir_f16_gpa (miss)
n_course (miss)
s1_female (miss)
Control
Treatment
1 2 3 4 5
2 3 4 5
0 1 2 3 4
0 50 100 150 200
0 1 NA

D Appendix D

Linear models are based on a number of assumptions: a linear relationship, homoscedasticity (homogeneity of variance), a lack of correlation between the independent variables, and normally-distributed residuals. Figure 5 summarizes tests of these assumptions for the initial student similarity perception model. All assumptions seem to hold for this model. Figure 6 shows the same tests for the end-of-term student similarity perception model.

Table 11 shows the two primary significant models, redone using a random train/test split (80% training, 20% testing). The treatment effect coefficient for initial student similarity perception is close to the coefficient in the first version of the model (β_1 = .202, 95% CI [.137, .267]), but the coefficient for end-of-term student similarity perception differs more (β_1 = .137, 95% CI [.055,.219]). Table 12 shows metrics to assess the performance of the two models on the training and testing datasets. The root mean squared errors and mean absolute errors are slightly larger for the testing data, which makes sense, as the model was fit to the training data but not the testing data. From these numbers, we can conclude that minimal overfitting in the model has occurred.

Table 11: Train replication model results

	Dependent variable:	
	Student sim. 1	Student sim. 2
	(1)	(2)
Treatment	.202***	.137***
	(.137,.267)	(.055,.219)
Class size	.0005	.001*
	(-.0003,.001)	(-0.00000,.002)
Student female	.102***	.067
	(.034,.171)	(-.019,.153)
CGPA	.018	.055*
	(-.034,.070)	(-.010,.120)
Year of study	.058***	.068***
	(.025,.092)	(.025,.110)
Constant	3.064***	2.986***
	(2.875,3.253)	(2.742,3.230)
Observations	1,725	1,395
R^2	.035	.023
Adjusted R^2	.032	.019

*p<0.1; **p<0.05; ***p<0.01

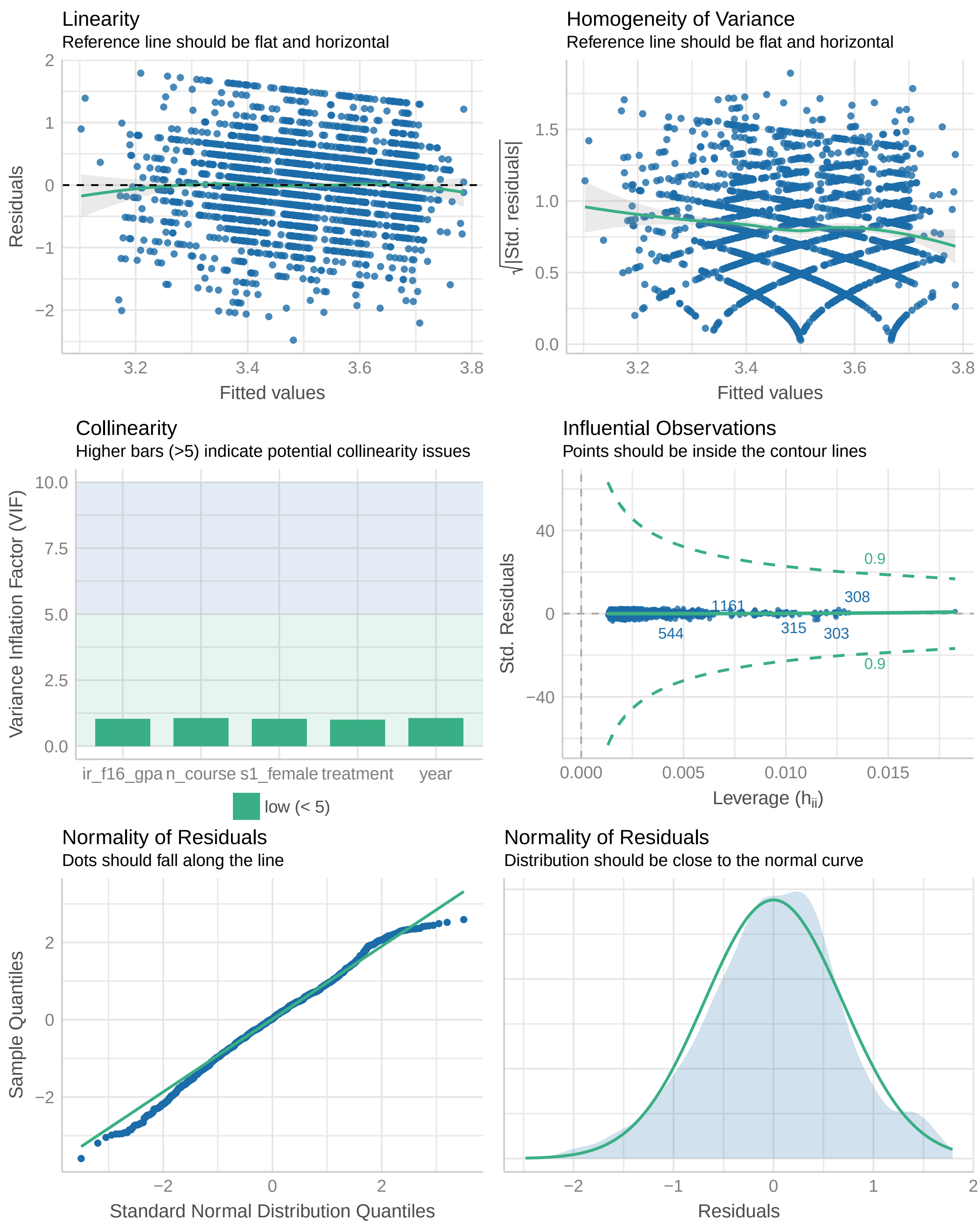


Figure 5: Initial student similarity perception model performance

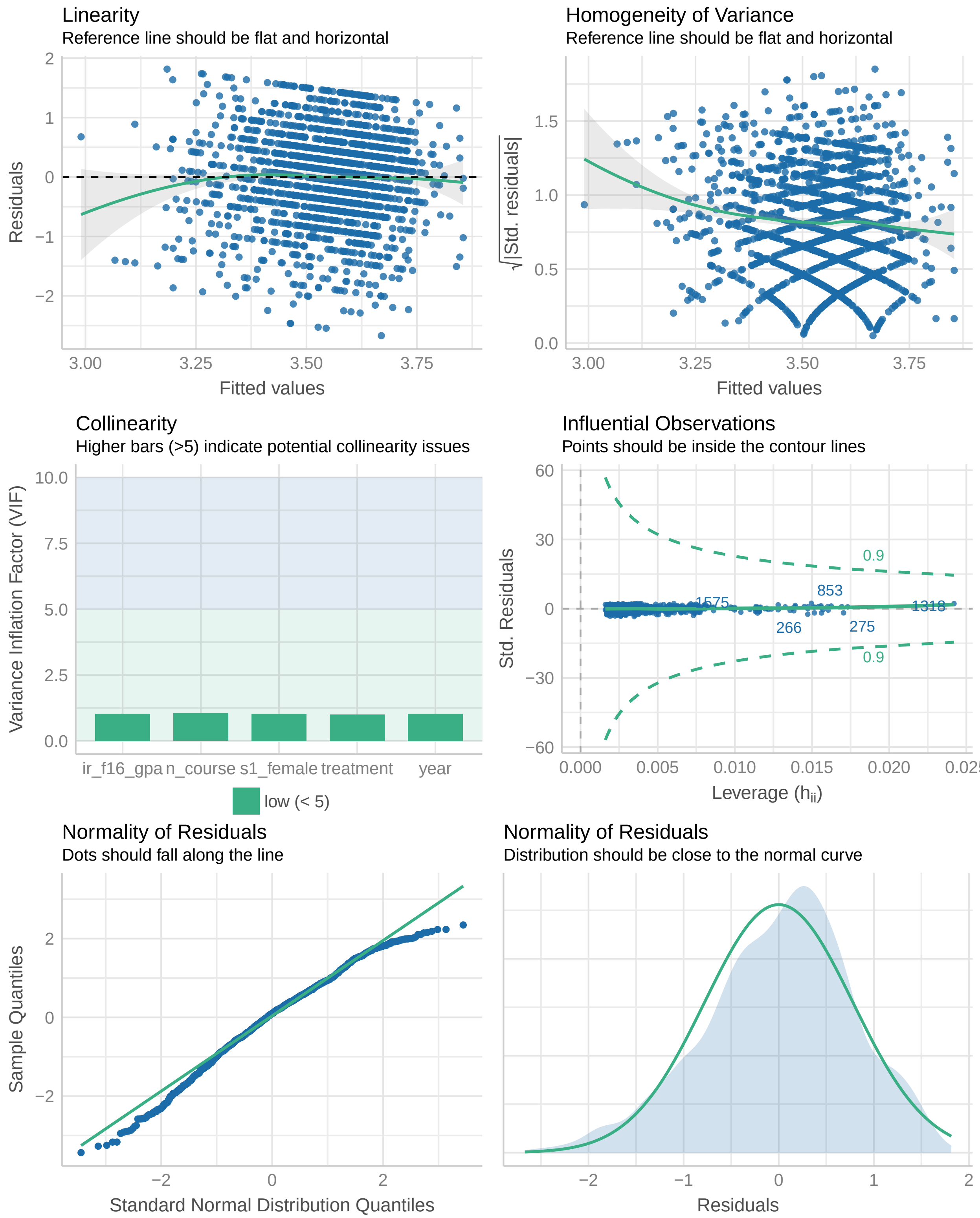


Figure 6: End-of-term student similarity perception model performance

Table 12: Metrics for student similarity perception models

Metric	Estimator	Training data	Testing data
Initial			
Root mean square error	standard	0.6884841	0.7079064
R-squared	standard	0.0352078	0.0285783
Mean absolute error	standard	0.5394814	0.5603628
End-of-term			
Root mean square error	standard	0.7766398	0.7926971
R-squared	standard	0.0225939	0.0256048
Mean absolute error	standard	0.6148309	0.6410355

E Appendix E

Table 13: Replication model results: Hispanic and Black students only

	Dependent variable:							
	Student sim. 1	Student sim. 2	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Instructor sim. 2	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*ordered logistic*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	$.198^{***}$	$.109^{**}$	$-.016$	$-.070$	$-.001$	$-.137$	$-.074$	.075
	(.122,.273)	(.013,.204)	(-.089,.057)	(-.162,.022)	(-.105,.103)	(-.410,.136)	(-.279,.130)	(-.499,.649)
Student ISR 1			$.673^{***}$					
			(.613,.733)					
Class size	.0002	.001	$-.0001$	$.002^{***}$	.0002	.0005	$.008^{***}$	$.009^{**}$
	(-.001,.001)	(-.0003,.002)	(-.001,.001)	(.001,.003)	(-.001,.001)	(-.002,.003)	(.006,.011)	(.00004,.017)
Student female	$.100^{**}$	.074	$-.021$	.043	.057	$-.153$	$.290^{**}$	$-.191$
	(.019,.181)	(-.030,.177)	(-.101,.059)	(-.056,.142)	(-.054,.169)	(-.469,.162)	(.068,.512)	(-.802,.419)
CGPA	$.057^{*}$	$.109^{***}$	.038	$.251^{***}$	$.678^{***}$	$.436^{***}$	$.561^{***}$	1.285^{***}
	(-.004,.117)	(.032,.186)	(-.021,.097)	(.177,.325)	(.595,.761)	(.248,.624)	(.391,.731)	(.903,1.666)
Year of study	$.044^{**}$	$.075^{***}$	$.080^{***}$	$.042^{*}$	$.139^{***}$	$.221^{***}$	$.102^{*}$	$-.123$
	(.007,.082)	(.027,.124)	(.043,.117)	(-.004,.088)	(.088,.190)	(.083,.359)	(-.002,.207)	(-.406,.160)
Constant	2.987^{***}	2.781^{***}	$.988^{***}$	2.462^{***}	$.513^{***}$	$.871^{***}$		$-.448$
	(2.775,3.199)	(2.507,3.054)	(.682,1.295)	(2.202,2.722)	(.224,.802)	(.217,1.525)		(-1.662,.766)
Observations	1,257	1,013	1,013	1,241	1,202	298	1,227	1,017
R^2	.034	.029	.346	.057	.215	.121		
Adjusted R^2	.030	.024	.342	.053	.211	.106		
Akaike Inf. Crit.								383.908

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

F Appendix F

Table 14: Replication model results: first-generation students only

	Dependent variable:							
	Student sim. 1	Student sim. 2	Student ISR 2	Instructor ISR 2	Course grade	Final grade	Instructor sim. 2	Persistence
	OLS	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*	*ordered logistic*	*logistic*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	.267***	.147***	−.018	−.048	−.023	−.104	.007	.101
	(.178,.355)	(.036,.258)	(-.104,.069)	(-.155,.059)	(-.142,.096)	(-.402,.195)	(-.229,.243)	(-.583,.784)
Student ISR 1			.662***					
			(.591,.733)					
Class size	.001	.001	−.0001	.003***	.0005	.001	.010***	.015**
	(-.0003,.001)	(-.0004,.002)	(-.001,.001)	(.001,.004)	(-.001,.002)	(-.001,.004)	(.007,.013)	(.002,.027)
Student female	.054	−.025	−.071	−.020	.033	−.010	.243*	−.957**
	(-.041,.149)	(-.144,.094)	(-.163,.022)	(-.135,.095)	(-.095,.160)	(-.365,.344)	(-.012,.497)	(-1.775,-.139)
CGPA	.050	.075*	−.002	.184***	.612***	.319***	.463***	1.314***
	(-.017,.117)	(-.008,.158)	(-.067,.062)	(.102,.265)	(.522,.702)	(.126,.512)	(.279,.647)	(.898,1.731)
Year of study	.037*	.063**	.069***	.066**	.159***	.289***	.110*	−.202
	(-.006,.081)	(.007,.120)	(.025,.113)	(.013,.120)	(.101,.217)	(.139,.440)	(-.009,.229)	(-.538,.134)
Constant	3.026***	2.971***	1.211***	2.592***	.659***	.896**		.091
	(2.789,3.263)	(2.669,3.274)	(.857,1.564)	(2.305,2.880)	(.343,.975)	(.204,1.587)		(-1.276,1.458)
Observations	949	770	770	936	910	220	925	768
R^2	.045	.022	.321	.052	.215	.141		
Adjusted R^2	.040	.016	.315	.047	.211	.121		
Akaike Inf. Crit.								271.574

*p<0.1; **p<0.05; ***p<0.01

G Appendix G

Table 15: Similarities presented to students and instructors during treatment

Similarities
Agreed! You both think instructors can best engage students by connecting the content to relevant issues.
Agreed! You both think that instructors can best engage students through interactive classes.
An instructor's most important characteristic? You both say a passion for their subject.
An instructor's most important characteristic? You both say being friendly/approachable.
An instructor's most important characteristic? You both say showing understanding.
An instructor's most important characteristic? You both say being knowledgeable.
Cool coincidence! You both think that instructors should tell students about the struggles they overcame to become a CSUN instructor.
Cool coincidence! You both think that instructors should tell students about their culture, personality, or hobbies.
Cool coincidence! You both think that instructors should tell students how they became interested in their subject.
Couch potatoes unite! When stressed you two like to binge watch TV.
Get to know the students! Instructors should know how students see themselves outside of class.
Get to know the students! Instructors should know their students' academic strengths.
Get to know the students! Instructors should know their students' academic weaknesses.
Get up on stage! You both enjoy public speaking to your peers.
Giving others the benefit of the doubt goes a long way. You both agree students and instructors should do this by assuming you both manage a lot outside of class.
Giving others the benefit of the doubt goes a long way. You both agree students and instructors should do this by learning what fears and anxieties you both face.
Giving others the benefit of the doubt goes a long way. You both agree students and instructors should do this by realizing you are both trying your best.
Good to know! You both want each other to know how hard you work for this class.
Good to know! You both want each other to know understand that this course matters, but there are other important things in your life.
Good to know! You both want each other to know what your other passions are outside of this topic.
Great minds prioritize alike! Establishing a work-life balance is a big deal for you both.
Great minds prioritize alike! Exploring who you are matters for you both.
Great minds prioritize alike! Finding a career connected to your passion is a big deal for you both.
Hire the same life coach? You both want to be more social.
Hire the same life coach? You both want to get to bed earlier.
Hire the same life coach? You both want to start exercising.
Hire the same life coach? You both want to stop procrastinating.
Hope you like the same teams! You two agree that it's fun to talk sports.
No sweat. When you get stressed, you both like to get a walk or some exercise.
Of all the challenges we face, you both agree that curbing terrorism needs to be a major focus.
Of all the challenges we face, you both agree that immigration reform needs to be a major focus.
Of all the challenges we face, you both agree that inequality needs to be a major focus.
Peas in a pod! You are both first generation college students.
Resilience matters–you both failed or almost failed a college class.
Resilience matters–you both have struggled in a college class.
Same New Year's Resolution? You both want to become more organized.
Same New Year's Resolution? You both want to better keep things in perspective.
Same New Year's Resolution? You both want to improve your communication skills.
Seems like you see eye-to-eye! Buckling down and studying for exams maximizes student learning.
Seems like you see eye-to-eye! Listening to interactive lectures maximizes student learning.
Seems like you see eye-to-eye! Working together in small groups maximizes student learning.

So long stress–you both mediate to de-stress.
So long stress–you both visit the snack bar to de-stress.
Solidarity! You both hate giving public speaking presentations to your peers.
Super important! You agree that instructors should learn about students' lives outside of class.
Super important! You agree that instructors should learn students' names.
Super important! You agree that instructors should learn whether students need help with the class material.
Support the students! You both want to build a better sense of community on campus.
Support the students! You both want to improve supports for student learning.
Support the students! You both want to reduce costs for students.
The door is open: You agree that professors should be there when students feel overwhelmed.
The door is open: You agree that professors should be there when students have questions about the course content.
The same concern keeps you two up at night: Figuring out future life decisions.
The same concern keeps you two up at night: Financial challenges.
The same concern keeps you two up at night: Getting motivated.
The two of you suspect instructors and students will connect better if they have faced similar challenges.
The two of you suspect instructors and students will connect better with a shared cultural background.
When times get tough for a student, you both agree that instructors should keep the same requirements for all students..
When times get tough for a student, you both agree that the instructor should be an understanding listener.
When times get tough for a student, you both agree that the instructor should give the student a second chance.
Whew – neither of you has ever been in danger of failing a class.
Woohoo! You both agree that instructors can best engage students by being enthusiastic.
You're on the same electronic page! Instructors and students should communicate via email.
You're on the same page! Instructors and students should communicate through one-on-one meetings.
You agree on the same secret to success: making mistakes is key!
You agree on the same secret to success: Students should build on their strengths!
You are cut from the same cloth–you both described yourself as caring!
You are cut from the same cloth–you both described yourself as outgoing!
You are cut from the same cloth–you both described yourself as thoughtful!
You are cut from the same cloth–you both described yourselves as funny!
You both agree that public speaking is not your favorite thing to do, but it's not the end of the world.
You both get it! Students learn best from applying classroom material to their personal lives.
You both get it! Students learn best through an instructors' individualized feedback.
You both get it! Students learn best from informal meetings with instructors.
You share a common goal: to foster a good instructor-student relationship.
You share a common goal: to see the subject matter differently by the end of the course.
You share a green perspective–solving the climate crisis is a top priority.
You share an outlook on friendship: friends should always have your back!
You share an outlook on friendship: friends should listen!
You share an outlook on friendship: friends should make you laugh!
You should connect? You agree that an instructor's career advice can be invaluable for students' careers.
You should connect? You agree that an instructor's help with professional networking can be invaluable for students' careers.
You should talk! You both agree that you should know what goals you share.
You should talk! You both agree that you should know what similar interests/passions you share.
You should talk! You two agree that it's fun to talk about TV, movies, etc..
You should talk! You two agree that it's fun to talk current events.
You two agree–both instructors and students should realize that a little understanding goes a long way.
You two agree–both instructors and students should realize that one class is just a small drop in the bucket.
You two agree–both instructors and students should realize that they both want the same goals.

References

Abrami, Philip C., and Deborah A. Mizener. 1985. "Student/Instructor Attitude Similarity, Student Ratings, and Course Performance." *Journal of Educational Psychology* 77 (6): 693–702. https://doi.org/10.1037/0022-0663.77.6.693.

Allaire, J J, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2020. "Rmarkdown: Dynamic Documents for R." https://github.com/rstudio/rmarkdown.

Bahns, Angela J., Christian S. Crandall, Omri Gillath, and Kristopher J. Preacher. 2017. "Similarity in Relationships as Niche Construction: Choice, Stability, and Influence Within Dyads in a Free Choice Environment." *Journal of Personality and Social Psychology* 112 (2): 329–55. https://doi.org/10.1037/pspp0000088.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2009. *College: What It Was, Is, and Should Be*. Princeton, U.S.A.: Princeton University Press.

Brinkworth, Maureen E., Joseph McIntyre, Anna D. Juraschek, and Hunter Gehlbach. 2018. "Teacher-Student Relationships: The Positives and Negatives of Assessing Both Perspectives." *Journal of Applied Developmental Psychology* 55: 24–38. https://doi.org/10.1016/j.appdev.2017.09.002.

Coolican, Hugh. 2014. *Research Methods and Statistics in Psychology*. 6th ed. Psychology Press. https://doi.org/10.4324/9780203769836.

Creasey, Gary, Patricia Jarvis, and Daniel Gadke. 2009. "Student Attachment Stances, Instructor Immediacy, and Student–Instructor Relationships as Predictors of Achievement Expectancies in College Students." *Journal of College Student Development* 50 (4): 353–72. https://doi.org/10.1353/csd.0.0082.

Delbanco, Andrew. 2015. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, U.S.A.: Princeton University Press.

Engineering, Stefan. 2018. "Create an Apa Style Correlation Table with R." *Blog Entry*. https://stefaneng.github.io/apa_correlation_table/; GitHub.

Evans, William N., Melissa S. Kearney, Brendan Perry, and James X. Sullivan. 2020. "Increasing Community College Completion Rates Among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case-Management Intervention." *Journal of Policy Analysis and Management* 39 (4): 930–65. https://doi.org/10.1002/pam.22256.

Garbuszus, Jan Marvin, and Sebastian Jeworutzki. 2018. *Readstata13: Import 'Stata' Data Files*. https://CRAN.R-project.org/package=readstata13.

Gehlbach, Hunter, Maureen E. Brinkworth, and Anna D. Harris. 2012. "Changes in Teacher–Student Relationships." *British Journal of Educational Psychology* 82 (4): 690–704. https://doi.org/10.1111/j.2044-8279.2011.02058.x.

Gehlbach, Hunter, Maureen Elizabeth Brinkworth, Aaron King, Laura Hsu, Joe McIntyre, and Todd T. Rogers. 2015. "Creating Birds of Similar Feathers: Leveraging Similarity to Improve Teacher-Student Relationships and Academic Achievement." https://doi.org/10.1037/edu0000042.

Harrell, Frank E. Jr., and Charles Dupont. 2020. *Hmisc: Harrell Miscellaneous*. https://CRAN.R-project.org/package=Hmisc.

Harrison, Ewen, Tom Drake, and Riinu Ots. 2020. *Finalfit: Quickly Create Elegant Regression Results Tables and Plots When Modelling*. https://CRAN.R-project.org/package=finalfit.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Iannone, Richard. 2020. *DiagrammeR: Graph/Network Visualization.* https://CRAN.R-project.org/package=DiagrammeR.

Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles.* https://www.tidymodels.org.

Lai, Tzu-Ling. 2015. "Effects of Student-Teacher Congruence on Students' Learning Performance: A Dyadic Approach." *Social Science Quarterly* 96 (5): 1424–35. https://doi.org/10.1111/ssqu.12215.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner, and Dominique Makowski. 2021. "Assessment, Testing and Comparison of Statistical Models Using R." *Journal of Open Source Software* 6 (59): 3112. https://doi.org/10.31234/osf.io/vtq8f.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reason, Robert Dean. 2009. "An Examination of Persistence Research Through the Lens of a Comprehensive Conceptual Framework." *Journal of College Student Development* 50 (6): 659–82. https://doi.org/10.1353/csd.0.0098.

Rizopoulos, Dimitris. 2006. "Ltm: An R Package for Latent Variable Modelling and Item Response Theory Analyses." *Journal of Statistical Software* 17 (5): 1–25. http://www.jstatsoft.org/v17/i05/.

Robinson, Carly D., Whitney Scott, and Michael A. Gottfried. 2019. "Taking It to the Next Level: A Field Experiment to Improve Instructor-Student Relationships in College." *AERA Open* 5 (1): 1–15. https://doi.org/10.21105/joss.01686.

———. 2020. "Instructor-Student Relationship Experiment: Data Package." *Inter-University Consortium for Political and Social Research [Distributor].* Inter-university Consortium for Political; Social Research (ICPSR). https://doi.org/10.3886/E119071V2.

Roorda, Debora, Helma MY Koomen, Jantine L. Spilt, and Frans J. Oort. 2011. "The Influence of Affective Teacher-Student Relationships on Students' School Engagement and Achievement: A Meta-Analytic Approach." https://doi.org/10.3102/0034654311421793.

Tinto, Vincent. 1997. "Classrooms as Communities: Exploring the Educational Character of Student Persistence." *The Journal of Higher Education* 68 (6): 599–623.

———. 2006. "Research and Practice of Student Retention: What Next?" *Journal of College Student Retention : Research, Theory & Practice* 8 (1): 1–19. https://doi.org/10.2190/4YNU-4TMB-22DJ-AN4W.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S.* Fourth. New York: Springer. http://www.stats.ox.ac.uk/pub/MASS4/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wilke, Claus O. 2020. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'.* https://CRAN.R-project.org/package=cowplot.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown.* https://github.com/rstudio/bookdown.

Zheng, Yuyan, Xiaotong Zheng, Chia-Huei Wu, Xiang Yao, and Yi Wang. 2021. "Newcomers' Relationship-Building Behavior, Mentor Information Sharing and Newcomer Adjustment: The Moderating Effects of Perceived Mentor and Newcomer Deep Similarity." *Journal of Vocational Behavior* 125: 103519. https://doi.org/10.1016/j.jvb.2020.103519.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.